

## Hoja de Trabajo No.2 Repaso Teoría

1. *¿Cuál es la principal innovación de la arquitectura Transformer?*
  - a. La principal innovación con la que cuenta la arquitectura Transformer es que este evita la recurrencia y se basa en un mecanismo de atención con la finalidad de dibujar dependencias globales en las entradas y salidas del modelo. Por otra parte, este permite la paralelización por lo que se puede alcanzar un nuevo nivel de precisión.
2. *¿Cómo funciona el mecanismo de atención del scaled dot-product?*
  - a. El mecanismo que tiene el scaled dot-product es que primero este recibe las consultas, claves y valores para luego calcular la similitud entre vectores por medio del cálculo del producto punto escalar entre las consultas y las claves. Seguidamente se realiza la función de atención con la finalidad de evitar obtener números grandes, esto se hace multiplicando la matriz de consultas y la transposición de las claves y dividiendo por la raíz cuadrada de la matriz de las claves para finalmente multiplicarlo por los valores, es decir  $((dq * tdk) / \sqrt{d_k}) dv$ . Por último a la función de activación se calcula la función softmax para así obtener los pesos de los valores.
3. *¿Por qué se utiliza la atención de múltiples cabezales en Transformer?*
  - a. La razón por la cual los Transformer hacen uso de múltiples cabezales radica en que estos le permiten al modelo atender un conjunto de datos que se encuentren en diferentes subespacios y posiciones. Es importante destacar que de esta manera el costo computacional es menor ya que se trabaja de forma paralela.
4. *¿Cómo se incorporan los positional encodings en el modelo Transformer?*
  - a. Esta consiste en agregar información de la posición a un vector, cabe destacar que estas tienen la misma dimensión que las representaciones del modelo buscando así que estas se puedan sumar. Esta hace uso de una función sinusoidal en donde cada dimensión de la codificación posicional corresponde a una senoide generando así una progresión geométrica de  $2\pi$  a 10000.
5. *¿Cuáles son algunas aplicaciones de la arquitectura Transformer más allá de la machine translation?*
  - a. Esta arquitectura también puede ser aplicada para generar de forma automática resúmenes a partir de porciones de texto relativamente grandes, por otra parte este se puede utilizar para encontrar y clasificar los diferentes elementos de un texto en categorías pre-definidas, entre otros.